International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

On the Design of Catastrophic Risk Portfolios

Ermoliev, Y.M., Ermolieva, T.Y., MacDonald, G.J. and Norkin, V.I.

IIASA Interim Report
August 1998

IIASA

International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-056 / August

On the design of catastrophic risk portfolios

Yuri M. Ermoliev (ermoliev@iiasa.ac.at)
Tatiana Y. Ermolieva (ermol@iiasa.ac.at)
Gordon MacDonald (macdon@iiasa.ac.at)
Vladimir I. Norkin (norkin@dept130.cyber.kiev.ua)

Approved by
Gordon MacDonald (macdon@iiasa.ac.at)
Director, *IIASA*

Abstract

Catastrophes produce rare and highly correlated insurance claims, which depend on the amount of coverage at different locations. A joint probability distribution of these claims is analytically intractable. The most promising approach for estimating total claims for a particular combination of decision variables involves geographically explicit simulations of catastrophes. The straightforward use of catastrophe models runs quickly into infinite "if – then" evaluations. The aim of this paper is to develop a framework allowing for the use of Monte Carlo simulation of catastrophes to aid decision making on designing optimal catastrophic risk portfolios. A dynamic stochastic optimization model is discussed. Connections between ruin probability and nonsmooth, in particular concave, risk functions are established. Nonsmooth adaptive Monte Carlo optimization is proposed.

Keywords: Catastrophes, Insurance, Risk, Stochastic optimization, Adaptive Monte Carlo, Nonsmooth optimization, Ruin probability.

Contents

1 Introduction **1**

2 Stochastic Optimization Model **2**
2.1 Risk Reserves 2
2.2 Model 3
2.3 Pareto Optimal Coverages 4

3 Probability of Ruin and Nonsmooth Risk Functions **5**

4 Nonsmooth Adaptive Monte Carlo Optimization **7**
4.1 Generalized Differentiability 7
4.2 Method 9

5 Numerical Experiments **10**

6 Concluding Remarks **11**

References **12**

On the design of catastrophic risk portfolios

Yuri M. Ermoliev (ermoliev@iiasa.ac.at)
Tatiana Y. Ermolieva (ermol@iiasa.ac.at)
Gordon MacDonald (macdon@iiasa.ac.at)
Vladimir I. Norkin (norkin@dept130.cyber.kiev.ua)

1 Introduction

Traditional insurance operates on the assumption of independent, frequent, low-consequence (conventional) risks, such as car accidents, for which decisions on premiums, estimates of claims and likelihood of insolvency (probability of ruin) can be calculated by using rich historical data. The law of large numbers provides in this case a simple "more-risk-is-better" portfolio selection strategy: if the number of independent risks in the portfolio is larger, then the variance of aggregate claims is lower and lower premiums can be chosen. This increases the demand for insurance, the coverage of losses, and, hence, the profits of insurers; it therefore also increases the stability of the insurance industry. The frequent occurance of conventional risks also permits simple "trial-and-error" or "learning-by-doing" procedures for adjusting default decision variables, for instance, premiums and coverage.

Traditional (collective) risk theory [1], [3], [4] relies on the law of large numbers, which allows for the pooling of data from multiple sources of loss in order to obtain collective estimated of the frequency parameters, aggregate losses and ruin probabilities. The emphasis on analytical approaches requires special assumptions on the underlying probability distribution. The ruin probability is usually analysed for infinite time horizons. The importance of dependencies in this analysis is discussed in [20]

Rare catastrophic risks require new portfolio selection approaches. Catastrophes produce claims highly correlated in space and time, which depend on the clustering of property and other values in the region and on geographical patterns of catastrophes, for example, natural disasters due to the persistence in climate [14]. The law of large numbers does not operate (in general) and the "more-risk-is-better" strategy may increase the probability of ruin for many insurers. The portfolio selection problem in the case of catastrophic risk is transformed from a purely statistical problem into a challenging risk selection problem. A principal difficulty is the lack of historical data on the occurrence of catastrophes at a particular location, although rich data may exist on their occurrence and magnitudes on a regional level. Potential losses at a particular location may be unlike any experienced in the past.

The most promising method for estimating dependent catastrophic losses for a particular combination of coverages and other decision variables involves the direct simulation of catastrophes, or catastrophe modeling [12]. This technique is becoming increasingly important to insurance companies as they make decisions on the allocation and values of contracts, premiums, reinsurance agreements, and the effects of mitigation measures. It is possible to simulate different patterns of catastrophes in a region realistically and to analyze the impact of different combinations of decision variables on the stability of

insurance companies. Unfortunately, this analysis runs quickly into infinite evaluations of "if-then" situations.

The aim of this paper is to develop a nonsmooth stochastic optimization techniques that allow the analyst to track spatial and temporal dependencies of losses and to direct adjustments of decision variables towards desirable outcomes by using Monte Carlo simulations. Catastrophes are extreme events and, as such, their analysis requires explicit introduction of nonsmooth (possibly discontinious) functions. Section 2 describes the dynamic stochastic optimization model, which is similar to those proposed in [6], [7]. This model extends classical results (Borch [2]) on the risk sharing to the case of non-substitutable risks, complex dynamics and possibility of the ruin. Section 3 establishes connections between nonsmooth risk functions and the probability of ruin. An adaptive Monte Carlo optimization procedure is analyzed in Section 4. Section 5 outlines numerical experiments, Section 6 presents some concluding remarks.

2 Stochastic Optimization Model

2.1 Risk Reserves

Assume that the study region is divided into subregions or locations $j = 1, 2, ..., m$. Locations may correspond to a collection of households, a zone with similar seismic activity, a watershed, etc. For each location j there exists an estimation W_j^t of the property value or "wealth" at time intervals $t = 0, 1, ...$, that includes values of houses, factories, etc. A sequence of random catastrophic events $\omega = \{\omega_t, t = 0, 1, ...\}$ affects different locations $j = 1, 2, .., m$ and generates at each $t = 0, 1, ...$ losses $L_j^t(\omega)$. These losses include direct losses from ω_t and indirect or delayed losses from previous time intervals. We assume that ω is an element of a probability space $(\Omega, \mathcal{F}, P)$, where Ω is a set of all possible ω, and $\mathcal{F}$ is a σ-algebra of measurable (with respect to probability measure P) events from Ω. We denote as $\{\mathcal{F}_t\}$ an increasing family of σ-algebras, $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$, $\mathcal{F}_t \subseteq \mathcal{F}$. Random variables $L_j^t(\omega)$ are assumed to be $\mathcal{F}_t$ - measurable, i.e., they depend on the observable "history" till t.

Losses $L_j^t(\omega)$, in contrast to conventional risks, are shared by many participants, such as governments, insurers, reinsurers, banks, and brokers. In the model these are called "insurers", although each participant shares a part of the risk and exhibits both insurer and reinsurer features.

For each insurer i the main variable of concern is his risk reserve R_i^t at time $t = 0, 1, ...$, or the money that the insurer has at its disposal:

$$R_i^{t+1} = R_i^t + \pi_i^t - C_i^t - S_i^t, t \geq 0,$$

where R_i^0 is a fixed amount of the initial risk reserve. At $t = 0, 1, ...$ premiums π_i^t push the trajectory of R_i^t up, whereas transaction costs C_i^t push it down. Claims S_i^t arriving at random moments trigger sudden jumps of R_i^t downwards.

In conventional risk theory the probability distribution of claim process S_i^t can be derived by using historical data. In the case of catastrophic risks there are strong dependencies among the variables S_i^t, $i = 1, ..., n$, which are affected by insurers' decisions on the spread of coverages among different locations. Since the joint probability distribution of claims is analytically intractable, we assume that there exists a Monte Carlo catastrophe model simulating trajectories of S_i^t, $t = 0, 1, ...$ Let us denote by q_{ij}^t a fraction of $L_j^t(\omega)$ covered by insurer i, i.e.,

$$\sum_{i=1}^{n} q_{ij}^t \leq 1, q_{ij}^t \geq 0. \tag{1}$$

Then the claim process can be written as

$$S_i^t(\omega) = \sum_{j \in I_i(t,\omega)} L_j^t q_{ij}^t, \tag{2}$$

where $I_i(t,\omega)$ is a subset of locations affected by ω till time t where insurer i still operates.

Remark 2.1 *In (2) a simple linear coverage function $q_{ij}^t(L_j^t) = q_{ij}^t L_j^t$ is used. There may be more general piece-wise linear coverage function $q_{ij}^t(L_j^t) = q_{ij}^{kt} L_j^t$, if $u_j^{k-1t} \le L_j^t \le u_j^{kt}$, $u_j^{k0} \ge 0$, $\sum_{i=1}^n q_{ij}^{kt} \le 1$, $q_{ij}^{kt} \ge 0$. According to this function different "slices" of losses L_j^t are covered with different fractions q_{ij}^{kt}. Parameters q_{ij}^{kt}, u_j^{kt} are decision variables; for example, $u_j^{0t} > 0$ indicates a deductible policy. The conventional reinsurance is associated with only two "slices" separated by an insurance "cap". In what follows only linear coverage functions are considered, although results hold for general functions.*

By using (2), risk reserve R_i^t is calculated as

$$R_i^{t+1} = R_i^t + \sum_{j=1}^m \left[\pi_{ij}^t(q^t) - C_{ij}^t(q^t)\right] - \sum_{j \in I_i(t,\omega)} L_j^t(\omega) q_{ij}^t, \tag{3}$$

where $i = 1, 2, ..., n$, $q^t = \{q_{ij}^t, i = \overline{1,n}, j = \overline{1,m}\}$, $t = 0, 1, ..., T-1$, and R_i^0 is an initial risk reserve.

2.2 Model

Without insurance, location j faces losses L_j^t. Individuals from this location receive compensation $L_j^t q_{ij}^t$ from company i when such a loss occurs. If W_j^0 is the initial wealth, then location j's wealth at time $t+1$ is

$$W_j^{t+1} = W_j^t + \sum_{i=1}^n \left(L_j^t q_{ij}^t - \pi_{ij}^t(q^t)\right) - L_j^t. \tag{4}$$

Individuals maximize their wealth, which depends on

$$v_j^t = \sum_{k=0}^{t-1} \left(L_j^t \sum_{i=1}^n q_{ij}^k - \sum_{i=1}^n \pi_{ij}^k(q^k) \right).$$

Therefore assume that coverages q_{ij}^t are chosen from the maximization of the expectation function

$$F_j(q) = \mathbf{E} f_j^{\tau_j}(x,\omega),\ f_j^{\tau_j} = v_j^{\tau_j - 1} + \gamma_j \min_{t \le \tau_j} \left[v_j^{t-1} - \mathbf{E} v_j^{t-1}\right] \tag{5}$$

subject to

$$\sum_{i=1}^n q_{ij}^t \le 1, j = \overline{1,m}, t = 0, 1, ..., T-1, \tag{6}$$

where γ_j is a substitution coefficient (or risk coefficient) between possible wealth and the risk of underestimating losses, τ_j is a stopping time, for example, the time of ruin not exceeding $T-1$, $\tau_j = \min\left[T-1, \min\left\{t : W_j^t \le 0, t \le T-1\right\}\right]$, $[a] = \min\{0, a\}$. Similarly, insurer i maximizes (by choosing coverages q_{ij}^t) his expected wealth

$$r_i^t = \sum_{k=0}^{t-1} \left\{ \sum_{j=1}^m \left[\pi_{ij}^k(q^k) - C_{ij}^k(q^k)\right] - \sum_{j \in I_i(t,\omega)} L_j^t(\omega) q_{ij}^t \right\},$$

taking into account the risk of overestimating profits and the risk of insolvency ($R_i^t < 0$). Coverages q_{ij}^t are chosen from maximization

$$G_i(q) = \mathbf{E}g_i^{\varphi_i}(t,\omega), g_i^{\varphi_i} = r_i^{\varphi_i - 1} + \varepsilon_i \left[\min_{t \le \varphi_i} r_i^{t-1} - \mathbf{E}r_i^{t-1}\right] + \delta_i \min\{0, R_i^{\varphi_i}\} \tag{7}$$

subject to (6), where ε_i, δ_i are substitution coefficients between profit and the risk of overestimating profits and insolvency, and φ_i is a stopping time, e.g.,

$$\varphi_i = \min\left[T-1, \min\left\{t : R_i^t \le 0, t \le T-1\right\}\right] .$$

Remark 2.2 *In the general case in (5), (7) can be used valuations $F_j(q) = \mathbf{E}f_j(W_j^t, 0 \le t \le \tau_j)$, $G_i(q) = \mathbf{E}g_i(R_i^t, 0 \le t \le \varphi_i)$ for some functions $f_j(\cdot)$, $g_i(\cdot)$. The maximization of (5) and (7) generates the insurance-demand functions and the insurance-supply functions depending on premiums. The choice of premiums must reflect balances between insurance demand and supply, otherwise higher premiums may decrease profits. In this paper we do not analyze the choice of premiums from this general perspective.*

2.3 Pareto Optimal Coverages

A Pareto optimal improvement of the initial catastrophic risk situation with respect to goal functions $F_j(q)$, $G_i(q)$ can be achieved by maximizing

$$W(q) = \sum_{j=1}^{m} \alpha_j F_j(q) + \sum_{i=1}^{n} \beta_i G_i(q), \tag{8}$$

subject to

$$\sum_{i=1}^{n} q_{ij}^t \le 1, q_{ij}^t \ge 0, j = 1,2,...,m, t = 1,2,...,T, \tag{9}$$

where $\alpha_j > 0$, $\beta_i > 0$, $\sum_{j=1}^m \alpha_j + \sum_{i=1}^n \beta_i = 1$. Let

$$W(q,\omega) = \sum_{j=1}^{m} \alpha_j f_j^{\tau_j}(q,\omega) + \sum_{i=1}^{n} \beta_i g_i^{\varphi_i}(q,\omega).$$

Then $W(q)$ can be written as $W(q) = \mathbf{E}W(q,\omega)$.

Random functions $W(q), W(q,\omega)$ have a complex analytical structure: they are, in fact, functionals of stochastic spatial processes (random fields) defined by simulated patterns of catastrophes. The nonsmooth character of functions $F_j(q)$ is due to the presence of operations min, max, and stopping times τ_j, φ_i in the definition of $W(q,\omega)$. This becomes more complex for general coverage functions.

Remark 2.3 *The above model can be modified for analyzing the capacity of the insurance "industry" in case of the most damaging catastrophic events. For this purpose [6] only uncertainties with sufficient historical data are characterized by random variables. Other uncertainties are considered from the worst-case perspective, consistent with their spatial patterns. For example, the occurrence of events in a region and their magnitudes can be characterized by a given probability distribution (Poisson, Pareto), whereas geographical location and their patterns can be chosen from the worst case. The resulting stochastic maximin model is a tradeoff between a conservative worst-case approach (all catastrophes are clustered at once in the most "valuable" locations) and the above model.*

Remark 2.4 *The use of stopping time arguments in problems (5)-(7) generally destroys the concavity of expectation $F_j(q)$, $G_i(q)$ despite the concavity of the components involved. These functions are concave when stopping times coincide with the moment of the first catastrophe. This important case reflects the nature of catastrophes as extreme events challenging the stability of the whole system once they occur.*

3 Probability of Ruin and Nonsmooth Risk Functions

There is a flexibility in choosing the weights α_j, β_j, ε_i, γ_j, δ_i. Coefficients α_j, β_j are responsible for the Pareto optimality. In (5), (7) nonsmooth risk functions are used to guarantee a trade-off between profits and risks of underestimating losses and overestimating profits with substitution coefficients ε_i, γ_j. These risk functions correspond to the Markovitz mean-semivariance model [15], the Konno and Yamazaki model [13] with absolute deviations, and the S. Messner et al. dynamic energy model [16]. In [19] it was shown that the use of absolute deviations with appropriate choice of risk coeeficients (similar to ε_i, γ_j) is consistent with the stochastic dominance of random outcomes. The applicability of the well-known mean-variance model [15] is usually linked with the normality of the probability distribution summarizing different prospects, which can not be assumed for catastrophic risks.

A key issue for catastrophic portfolio selection problems is the possible ruin of insurers. Let us show that when risk coefficients δ_i become large enough, then the probability of ruin drops below a given level.

The function $W(q)$ can be represented in the form

$$W(q) = V(q) + \mathbf{E}\sum_{i=1}^{n} \beta_i \delta_i \min\left\{0, R_i^{\varphi i}\right\} \ .$$

If $\delta_i = N/\beta_i$, where N is a large number, then $W(q) = V(q) + N\mathbf{E}\sum_{i=1}^{n} \min\{0, R_i^{\varphi_i}\}$. Let us show that if N is large enough, then maximization of $W(q)$ approximates the maximization of $V(q)$ subject to the chance constraints $\mathbf{P}\left\{\sum_{i=1}^{n} \min\{0, R_i^{\varphi_i}\} < 0\right\} < \varepsilon$ for arbitrary small $\varepsilon > 0$; that is, the ruin probability of any insurer cannot fall below a given level. This is due to the following general result, which for the case of linear chance constraints was, in fact, discussed in [22].

Consider two problems, the chance constraint problem

$$F(x) \longrightarrow \max_{x \in X} \tag{10}$$

subject to

$$P(x) = \mathbf{P}\{g(x, \omega) > 0\} \leq \epsilon, \tag{11}$$

with optimal value F_ϵ^* and the problem

$$\Phi_N(x) = F(x) + NG_+(x) \longrightarrow \max_{x \in X}, \tag{12}$$

with optimal value Φ_N^*, where $X \subset R^n$ is a compact set, $F(x)$ is a continuous function , $G_+(x) = \mathbf{E}\max\{0, g(x, \omega)\}$, and N is a penalty coefficient. Here ω denotes an elementary event in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Assume that

(i) $g(\cdot, \omega)$ is almost sure (a.s.) continuous and $|g(x, \omega)| \leq C(\omega)$ for all $x \in X$, $\mathbf{E}C^{1+\lambda}(\omega) \leq C^{1+\lambda} < +\infty$, for some $C > 0$, $\lambda > 0$.

(ii) $G_+(x') = 0$ for some $x' \in X$;

(iii)

$$\mathbf{P}\{g(x,\omega)=0\}=0 \ \ \forall\, x\in X. \tag{13}$$

Assumption (iii) implies that function $P(x)=\mathbf{P}\{g(x,\omega)>0\}$ in (11) is continuous (see [21]). The following lemma shows that if penalty term $G_+(x_N)$ goes to zero for some sequence of points $\{x_N\}$ then $P(x_N)$ also go to zero as $N\longrightarrow+\infty$.

Lemma 3.1 *Let for some sequence of points* $\{x_N\}$ $\lim_{N\to+\infty}G_+(x_N)=0$. *Then*

$$\lim_{N\to+\infty}\mathbf{P}\{g(x_N,\omega)>0\}=0.$$

Proof 3.1 *Denote* $\delta_N=G_+(x_N)$ *and* $\mathcal{F}_x(t)=\mathbf{P}\{g(x,\omega)\le t\}$. *By Chebyshev inequality:*

$$\begin{aligned}\mathbf{P}\{g(x_N,\omega)>0\} &= \mathbf{P}\{0<g(x_N,\omega)\le\sqrt{\delta_N}\}+\mathbf{P}\{g(x_N,\omega)>\sqrt{\delta_N}\}\\ &\le \mathcal{F}_{x_N}(\sqrt{\delta_N})-\mathcal{F}_{x_N}(0)+\mathbf{P}\{\max(0,g(x_N,\omega))>\sqrt{\delta_N}\}\\ &\le \mathcal{F}_{x_N}(\sqrt{\delta_N})-\mathcal{F}_{x_N}(0)+\tfrac{1}{\sqrt{\delta_N}}\mathbf{E}\max(0,g(x_N,\omega))\\ &= \mathcal{F}_{x_N}(\sqrt{\delta_N})-\mathcal{F}_{x_N}(0)+\sqrt{\delta_N}.\end{aligned}$$

By condition (iii) the distribution function $\mathcal{F}_x(t)$ *is continuous at any point* $(x,0)$. *Without loss of generality we can assume that* $x_N\longrightarrow x$ *as* $N\longrightarrow+\infty$. *Since* $\delta_N\longrightarrow 0$, *from the continuity of* $\mathcal{F}_x(t)$ *at* $(x,0)$ *follows* $\mathcal{F}_{x_N}(\sqrt{\delta_N})\longrightarrow\mathcal{F}_x(0)$ *and* $\mathcal{F}_{x_N}(0)\longrightarrow\mathcal{F}_x(0)$. *Hence* $\mathbf{P}\{g(x_N,\omega)>0\}\longrightarrow 0$ *as* $N\longrightarrow+\infty$.

Lemma 3.2 *Let us assume that for any* $\epsilon>0$ *there exists a point* $x_\epsilon\in X$ *such that*

$$\mathbf{P}\{g(x_\epsilon,\omega)>0\}\le\epsilon.$$

Then

$$G_+(x_\epsilon)\le C\epsilon^{\lambda/(1+\lambda)} \tag{14}$$

and hence $\lim_{\epsilon\to0}G_+(x_\epsilon)=0$.

Proof 3.2 *Denote*

$$I_{g(x,\omega)>0}=\begin{cases}1, & g(x,\omega)>0,\\ 0, & \textit{otherwise.}\end{cases}$$

By Hölder inequality

$$\mathbf{E}\max(0,g(x_\epsilon,\omega)=\int_\Omega|g(x_\epsilon,\omega)|I_{g(x_\epsilon,\omega)>0}\mathbf{P}(d\omega)\le$$

$$\int_\Omega C(\omega)I_{g(x_\epsilon,\omega)>0}\mathbf{P}(d\omega)\le\left(\int_\Omega C^{1+\lambda}(\omega)\mathbf{P}(d\omega)\right)^{1/(1+\lambda)}\left(\int_\Omega I_{g(x_\epsilon,\omega)>0}\mathbf{P}(d\omega)\right)^{\lambda/(1+\lambda)}\le$$

$$C\left(\mathbf{P}\{g(x_\epsilon,\omega)>0\}\right)^{\lambda/(1+\lambda)}\le C\epsilon^{\lambda/(1+\lambda)}.$$

Thus, $\lim_{\epsilon\to0}G_+(x_\epsilon)=0$.

The next theorem relates the optimal values of the chance constraint problem (10), (11) and problem (12).

Theorem 3.1 *There exist non-negative functions* $\epsilon(N)$, $\alpha(N)$, $\beta(\epsilon)$ *and* $\gamma>0$ *such that*

$$\lim_{N\to+\infty}\epsilon(N)=\lim_{N\to+\infty}\alpha(N)=\lim_{\epsilon\to0}\beta(\epsilon)=0,$$

$$\Phi^*_N-\alpha(N)\le F^*_{\epsilon(N)}\le\Phi^*_{1/\epsilon^\gamma(N)}-\beta(\epsilon(N)), \tag{15}$$

$$F^*_{1/N^{1/\gamma}}+\beta(1/N^{1/\gamma})\le\Phi^*_N\le F^*_{\epsilon(N)}+\alpha(N). \tag{16}$$

Proof of Theorem 3.1 Denote x_ϵ, x_N optimal solutions of problems (10), (11) and (12) respectively, $\alpha(N) = NG_+(x_N)$, $\beta(\epsilon) = \epsilon^{-\frac{\lambda}{2(1+\lambda)}} G_+(x_\epsilon)$. $\epsilon(N) = \mathbf{P}\{g(x_N,\omega) > 0\}$. Then $\lim_{N\to+\infty} \alpha(N) = 0$ by properties of the penalty function method (see, for example, [10], [11]). By estimate (14) $\lim_{\epsilon\to 0} \beta(\epsilon) = 0$. From Lemma 3.1 it follows that $\lim_{N\to+\infty} \epsilon(N) = 0$. Since by definition $\mathbf{P}\{g(x_N,\omega) > 0\} \le \epsilon(N)$, then by optimality of x_ϵ

$$\Phi^*_N = F(x_N) + NG_-(x_N) \le F(x_{\epsilon(N)}) + NG_-(x_N) = F^*_{\epsilon(N)} + \alpha(N). \tag{17}$$

Denote $\gamma = \frac{\lambda}{2(1+\lambda)}$. By optimality of $x_{N(\epsilon)}$, $N(\epsilon) = \epsilon^{-\gamma}$

$$F^*_\epsilon = (F(x_\epsilon) + \epsilon^{-\gamma} G_-(x_\epsilon)) - \beta(\epsilon) \le \Phi^*_{\epsilon^{-\gamma}} - \beta(\epsilon). \tag{18}$$

Now fix an arbitrary $N > 0$. From (17) and (18) (with $\epsilon(N)$ instead of ϵ) follows (15). From (18) with $\epsilon = 1/N^{1/\gamma}$ and (17) follows (16).

Let us now come back to $W(q) = V(q) + N\sum_{i=1}^n \min\{0, R_i^{\varphi_i}\}$. If we use functions $F(x) := V(q)$, $g(x,\omega) := -\sum_{i=1}^n \min\{0, R_i^{\varphi_i}\}$ in (10), (11), then lemmas 3.1, 3.2 and theorem 3.1 show that the maximization of $W(q)$ for a large N indeed approximates the maximization of $V(t)$ subject to the ruin probability constraint.

4 Nonsmooth Adaptive Monte Carlo Optimization

4.1 Generalized Differentiability

Problems (5) and (7) have the following general structure. Let $\{V^t(x,\omega),\ 0 \le t \le T-1\}$ be a real-valued discrete time random (risk) process depending on deterministic vector parameter $x \in X \subset R^n$ and random parameter ω. Define a stopping time

$$\tau(x,\omega) = \min\left[T-1, \min\{t:\ V^t(x,\omega) < 0,\ 0 \le t \le T-1\}\right].$$

Consider a risk function $F^t(x) = \mathbf{E} f^t(x,\omega)$,

$$f^t(x,\omega) = \min_{0\le i<t} V^i(x,\omega) + \gamma_t(V^0,...,V^t),$$

where $\gamma_t(\cdot)$ is a nonsmooth function, and $F(x) = \mathbf{E} f(x,\omega)$

$$f(x,\omega) = \min_{0\le t<\tau(x,\omega)} V^t + \gamma_t(V^0,...,V^t)|_{t=\tau(x,\omega)}.$$

If functions $V^i(x,\omega)$, $\gamma_t(\cdot)$ are concave in x then $F^t(x)$ is also concave, but this is not the case with function $F(x)$ due to the dependence of $\tau(x,\omega)$ on x. Let us show that $F(x)$ is a generalized differentiable (GD) function assuming generalised differentiability of $V^t(x,\omega)$, $0 \le t \le T-1$, $\gamma_t(\cdot)$. The class of GD-functions is especially important for problems with general coverage functions, involving deductable and reinsurance "caps".

Definition 4.1 *[18] Function $f : R^n \longrightarrow R$ is called generalized differentiable at $x \in R^n$ if in some vicinity of x there exists an upper semicontinuous at x multivalued mapping ∂f with closed convex compact values $\partial f(x)$ such that*

$$f(y) = f(x) + \langle g, y-x\rangle + o(x,y,g), \tag{19}$$

where $\langle\cdot,\cdot\rangle$ denotes an inner product of two vectors, $g \in \partial f(y)$ and

$$\lim_k \frac{|o(x,y^k,g^k|}{\|y^k - x\|} = 0 \tag{20}$$

for any sequences $y^k \longrightarrow x$, $g^k \in \partial f(y^k)$. Function f is called generalized differentiable if it is generalized differentiable at each point $x \in R^n$.

The GD-functions possess the following properties ([17],[18]): they are continuously differentiable almost everywhere in R^n; and $\partial f(x)$ is a singleton for almost all $x \in R^n$. GD-functions are locally Lipschitzian, for Clark subdifferential $\partial_C f(x) \subseteq \partial f(x)$; continuously differentiable, convex and concave functions are generalized differentiable; class GD-functions are closed with respect to *max*, *min* operations and superpositions; there is a calculus of subgradients:

$$\partial \min(f_1, f_2)(x) = co\{\partial f_i | \ \ f_i(x) = \min(f_1(x), f_2(x))\}, \tag{21}$$

where $co\{\cdot\}$ denotes a convex hull of $\{\cdot\}$ and the subdifferential $\partial f_0(f_1, \ldots, f_m)$ of a composite function $f_0(f_1, \ldots, f_m)$, where $f_0(\cdot)$ is a GD-function, is calculated by the chain rule.

In addition, the class of GD-functions is closed with respect to taking expectations.

Theorem 4.1 *([17]). Let $(\Omega, \Sigma, \mathbf{P})$ be a probability space, function $f : R^n \times \Omega \longrightarrow R^1$ is generalized differentiable at $\overline{x} \in R^n$ for almost all $\omega \in \Omega$ and integrable in ω for all $x \in R^n$. Assume that gradient (in x) mapping $\partial f(x, \omega)$ is measurable in ω for all x (such mapping exists and can be constructed [17]), and for any compact $X \subset R^n$ there exists an integrable function $L_X(\omega)$, such that*

$$\sup\{|f(x, \omega)| \mid x \in X\} \leq L_X(\omega),$$

$$\sup\{\|g\| \mid g \in \partial f(x, \omega), \ x \in X\} \leq L_X(\omega).$$

Then $F(x) = \mathbf{E} f(x, \omega)$ is generalized differentiable at $\overline{x}$ with $\partial F(x) = \mathbf{E}\partial f(x, \omega)$.

Assuming that $V^t(x, \omega), \ t = 0, 1, \ldots, T-1$, are generalized differentiable functions, the above properties imply that $f^t(x, \omega)$ and (under appropriate assumptions) $F^t(x)$ are also generalized differentiable functions. The same is not so evident for $f(x, \omega)$ and $F(x)$ because $\tau(x, \omega)$ depends on $(x, \ \omega)$. The following theorem shows that under practically important conditions $F(x)$ is also a GD- function with a quite natural calculus of subgradients.

Theorem 4.2 *Assume that*

(i) functions $V^t(x, \omega), \ \gamma_t(\cdot), \ 0 \leq t \leq T$, are genralized differentiable in $x \in X$ for almost all ω and

$$\sup\{|V^t(x, \omega)| \ | x \in X\} \leq L(\omega), \ |\gamma_t(\cdot)| \leq L(\omega)$$

with integrable function $L(\omega)$, $t = 0, 1, \ldots, T$,

(ii) generalized gradient (in x) mappings $\partial V^t(x, \omega)$, $\partial \gamma_t$ are measurable in ω and bounded by $L(\omega)$ for all $x \in X$

(iii) for all $x \in X$ and $t = 0, 1, \ldots, T$, the probability

$$\mathbf{P}\{V^t(x, \omega) = 0\} = 0.$$

Then

(a) function $f(x, \omega)$ is a.s. generalized differentiable with

$$\partial f(x, \omega) = co\{\partial V^t | \ t \in t^*(x, \omega)\} + \partial \gamma_t(V^0, ..., V^t)|_{t=\tau(x,\omega)} \tag{22}$$

$$t^*(x, \omega) = \{t | \ V^t(x, \omega) = \min_{0 \leq t < \tau(x,\omega)} V^t(x, \omega), \ \ 0 \leq t < \tau(x, \omega)\},$$

(b) expectation function $F(x) = \mathbf{E} f(x, \omega)$ is generalized differentiable with

$$\partial F(x) = \mathbf{E}\partial f(x, \omega).$$

Proof of Theorem 4.2. Let us fix a point $\overline{x} \in X$. Denote $\overline{\Omega}$ the set of ω such that $V^t(\overline{x},\omega) = 0$ for at least one t, $0 \le t \le T-1$. By condition (iii) $\mathbf{P}\{\overline{\Omega}\} = 0$. Fix an abitrary $\omega \in (\Omega \setminus \overline{\Omega})$ and denote $\overline{\tau}(\omega) = \tau(\overline{x},\omega)$. Then $V^t(\overline{x},\omega) > 0$, $0 \le t < \overline{\tau}(\omega)$, $V^{\overline{\tau}(\omega)}(\overline{x},\omega) < 0$ or $V^t(\overline{x},\omega) > 0$, $0 \le t \le T-1$. By continuity of $V^t(x,\omega)$ there is a vicinity $\overline{U}(\omega)$ of $\overline{x}$ such that for $x \in \overline{U}(\omega)$ the following holds: $V^t(x,\omega) > 0$, $0 \le t < \overline{\tau}(\omega)$, $V^{\overline{\tau}(\omega)}(x,\omega) < 0$ or $V^t(x,\omega) > 0$, $0 \le t \le T-1$. Thus $\tau(x,\omega) = \overline{\tau}(\omega)$ and $f(x,\omega) = \overline{f}(x,\omega)$ for $x \in \overline{U}(\omega)$, where $\overline{f}(x,\omega) = \min_{0\le t<\overline{\tau}(\omega)} V^t + \gamma_t|_{t=\overline{\tau}(x)}$. Function $\overline{f}(x,\omega)$ is generalized differentiable at $\overline{x}$ with (upper semicontinuous) subdifferential

$$\partial\overline{f}(x,\omega) = co\{\partial V^t | t \in \overline{t}(x,\omega)\} + \partial\gamma_t|_{t=\overline{\tau}(\omega)},$$

$$\overline{t}(x,\omega) = \{t|\ V^t(x,\omega) = \min_{0\le t<\overline{\tau}(\omega)} V^t(x,\omega),\ \ 0 \le t < \overline{\tau}(\omega)\}.$$

Since $\partial f(x,\omega)$ given by (22) coincides with $\partial\overline{f}(x,\omega)$ in a vicinity of $\overline{x}$, $f(x,\omega)$ is generalized differentiable at $\overline{x}$ with subdifferential (22).

To prove statement (b) we check the conditions of Theorem 4.1. Obviously,

$$\sup\{|f(x,\omega)|\ |\ x \in X\} \le 2L(\omega),$$

$$\sup\{\|g\||\ g \in \partial f(x,\omega),\ \ x \in X\} \le 2L(\omega).$$

The stopping time $\tau(x,\omega)$ is a measurable in ω function, since

$$\{\omega|\ \tau(x,\omega) \le t < T\} = \{\omega|\ \min_{0\le i\le t} V^i(x,\omega) < 0\},$$

$$\{\omega|\ \tau(x,\omega) = T\} = \{\omega|\ \min_{0\le t<T} V^t(x,\omega) \ge 0,\ V^T(x,\omega) < 0\} \cup \{\omega|\ \min_{0\le t\le T} V^t(x,\omega) \ge 0\},$$

and for any t multivalued mapping $\partial f^t(x,\omega)$ is measurable; i.e., for any compact $K \subset R^n$ the set

$$\{\omega|\ \partial f^t(x,\omega) \cap K \ne \emptyset\}$$

is measurable, which can be easily demonstrated. Then multifunction $\partial f(x,\omega)$ is measurable, since for any compact $K \subset R^n$ the set

$$\{\omega|\ \partial f(x,\omega) \cap K \ne \emptyset\} = \cup_{0\le t\le T}\Big(\{\omega|\ \tau(x,\omega) = t\} \cap \{\omega|\ \partial f^t(x,\omega) \cap K \ne \emptyset\}\Big)$$

is measurable. Thus, the multifunction $\partial f(x,\omega)$ is convex, compact valued, and measurable. By (a) for any fixed $\overline{x}$ function $f(x,\omega)$ is almost sure generalized differentiable at $\overline{x}$ with subdifferential (22). Then statement (b) follows from Theorem 4.2.

4.2 Method

Maximization of $W(q)$ for general coverage functions (see Remark in section 2) and explicit insolvency constraints leads to a general nonsmooth stochastic optimization problem of the type: maximize $F(x) = \mathbf{E}f(x,\omega)$, $x \in X \subseteq R^l$ with GD-functions $F(\cdot)$, $f(\cdot,\omega)$, and $X = \{x|\Psi(x) \le 0\}$ defined by a GD-function $\Psi(x)$. Assume the following regularity condition: $\inf\{\|g\| : g \in \partial\Psi(x)\} > 0$, where $\partial\Psi$ is defined according to (4.1). The following key result was proved in [8]. Consider the stochastic quasigradient (SQG) procedure:

$$x^{k+1} \in \Pi_x(x^k - \rho_k\xi^k), x^0 \in X, \tag{23}$$

where x^k, ξ^k, $k = 0, 1, ...$ are defined on a probability space $(\Omega, \mathcal{F}, P)$,

$$\mathbf{E}\left\{\xi^k | x^0, ..., x^k\right\} \in \partial f(x^k,\omega),$$

and Π_x is a (multivalued) projection operator on the set X; $\rho_k \geq 0$, $\sum_{k=0}^{\infty} \rho_k = \infty$, $\sum_{k=0}^{\infty} \rho_k^2 < \infty$. Define $X^* = \{x | 0 \in \partial F(x) + N_X(x)\}$, where $N_X(x) = \{\lambda \partial \Psi(x) : \lambda \geq 0\}$ if $\Psi(x) = 0$ and $N_X(x) = 0$ if $\Psi(x) < 0$. Let X be a compact and $\left\|\xi^k(\omega)\right\| \leq C < \infty$(which usually follows from the compactness of X).

Theorem 4.3 *All cluster points of* $\left\{F(x^k)\right\}$ *a.s. constitute an interval in* F^*. *If set* F^* *does not contain intervals (for example,* F^* *is a finite or countable), then all cluster points of* $\left\{x^k(\omega)\right\}$ *a.s. belong to a connected subset of* X^* *and* $\left\{F(x^k(\omega)\right\}$ *has a limit in* F^*.

Theorems 4.2, 4.3 allow us to develop adaptive Monte Carlo optimization for rather general catastrophic risk selection problems. Assume that after k simulations of catastrophes $\omega(0), \omega(1), ..., \omega(k-1)$ a set of coverages $q(k) = \left\{q_{ij}^t(k), i = \overline{1,n}, j = \overline{1,m}, t = \overline{0,T-1}\right\}$ is obtained. Coverages $q(k)$ correspond to approximate solutions x^k, $k = 0, 1, ...$ in (23). From (22) follows a simple rule for calculating ξ^k: for given $q(k)$ simulate a new independent sequence of catastrophes $\omega(k) = (\omega^0(k), \omega^1(k), ..., \omega^{T-1}(k))$, observe stopping times $\tau_j^k = \tau_j(q(k), \omega(k))$, $\tau_i^k = \tau_i(q(k), \omega(k))$, and calculate subgradients of functions $f_j^t(q, \omega(k))$, $g_i^t(q, \omega(k))$ with respect to $q^t(k)$, $t \leq \tau_j^k$ and correspondingly $t \leq \tau_i^k$. Compute

$$\xi^k = \sum_{j=1}^{m} \alpha_j f_{jq}^t \left(q(k), \omega(k)\right)|_{t=\tau_j^k} + \sum_{i=1}^{n} \beta_i g_{iq}^t \left(q(k), \omega(k)\right)|_{t=\tau_i^k}.$$

After that a new set of coverages $q(k+1)$ is adjusted from $q(k)$ according to (23), etc., where the projection on the set defined by (6) is splited into T independent subproblems for each group of variables q^t, $t = 0, 1, ..., T-1$.

5 Numerical Experiments

Numerous numerical experiments on design of catastrophic risk insurance portfolios using the proposed approach are described in [6], [7]. They show a satisfactory speed of major improvements of initial coverages. Thus, Figure 1 illustrates typical dynamics of improvements during iterations $k = 0, 1, ...$ In this example the number of locations $m = 100$, insurance companies $n = 5$, the time span $T = 1000$, the stopping time coincides with the time of the first catastrophe, coverage functions are linear and do not depend on t. The indicator of improvements is the sample mean of dependent variables $W(q(k), \omega^k)$, $k = 0, 1, ...$, $W^k = (1/k) \sum_{s=0}^{k-1} W(q(s), \omega^s)$. It is possible to show [9] that the law of large numbers holds for this type of indicators and W^k approaches $W(q(k))$ when $k \to \infty$. Figure 1 shows the adaptation of initial coverages to catastrophes. As we can see, the initial coverages are sensitive to catastrophes. The sequential adaptive adjustments (23) improve their spatial diversification, which increases the "welfare" functions $W(q)$, i.e., provides Pareto optimal improvements with respect to the profits of insurers, losses of individuals and insurer insolvency. Numerical experiments in [6],[7] show that desirable histograms of insolvency can easily be achieved by simple "manipulations" with risk coefficients δ_i.

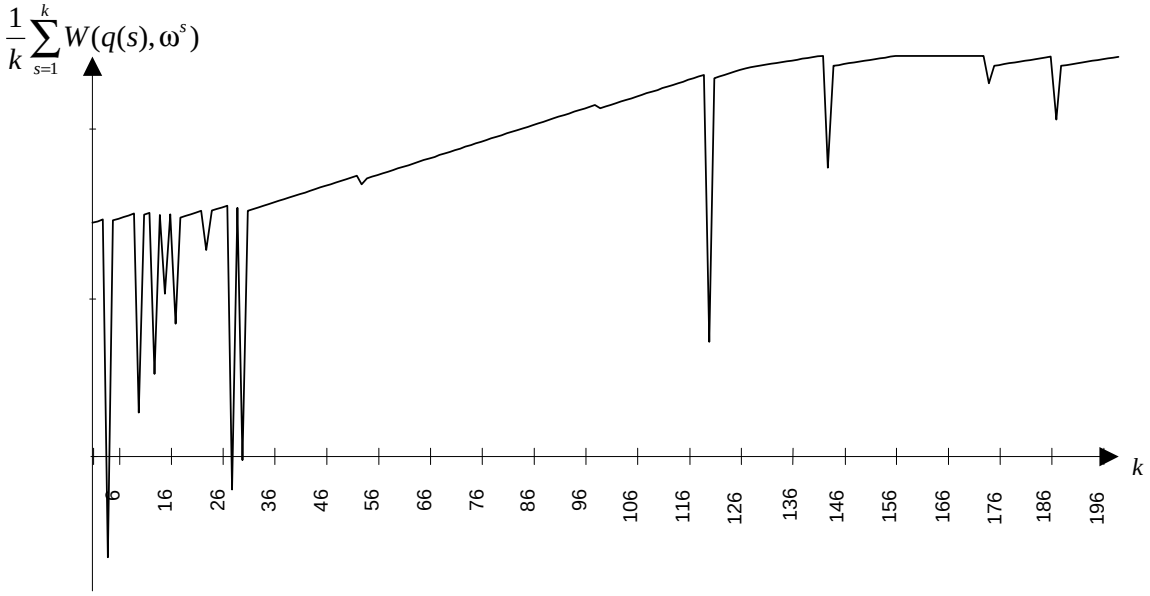

Figure 1: Improvements of performance indicator $\frac{1}{k}\sum_{S=1}^{K} W(q(s), w^s)$

6 Concluding Remarks

A key feature of a catastrophe risk selection problem is the insolvency (probability of ruin) of insurers. In this paper the insolvency is taken into account by a nonsmooth risk function. It leads to concave stochastic optimization problems in the case of a concave with respect to q functions $R_i^t(q,\omega)$, $W_j^t(q,\omega)$ and stopping times independent of q. In contrast, explicit introduction of a constraint on the probability of ruin destroys the concavity. Optimal selection of catastrophic risks for models of present complexity with stopping times can not be fully studied by analytical techniques and deterministic sample mean approximations. Therefore we use adaptive Monte Carlo optimization. Specific stochastic quasigradient methods enable us to deal with nonsmooth risk functions and implicit dependencies of stopping times on decision variables. Theorem 4.3 establishes the use of common random numbers resulting in a considerable increase of computational efficiency. Combination of proposed methods with other approaches and the variance-reduction techniques require special attention. The efficiency of the approach presented requires also the development of dynamic catastrophe models incorporating key variables responsible for the random occurance of specific catastrophes and dependencies (see discussion in [14] concerning natural catastrophes). This approach can be extended to more general problems with nonlinear coverage functions and other insurance-related variables, since the class of GD-functions is rich enough to model the nonsmooth character of various risk management situations (see discussion in [5]).

References

[1] R.Beard, T.Pentikainen, E.Pesonen, *Risk Theory*, Printed in Great Britain at the University Printing House, Cambridge, 1984.

[2] K.Borch, Equilibrium in a reinsurance market, Econometrica. Vol.30, 3 (1962).

[3] H.Buhlmann, *Mathematical Methods in Risk Theory*, Springer-Verlag, New York, Heidelberg Berlin, 1970.

[4] C.D.Daykin, T.Pentikainen and M.Pesonen, *Practical Risk theory for Actuaries*, Monographs on statistics and applied probability, vol. 53, Chapman and Hall Ltd., 1994.

[5] Y.Ermoliev and V. Norkin, On nonsmooth and discontinuous problems of stochastic systems optimization, European Journal of Operation Research, 1997.

[6] T.Ermolieva, The Design of Optimal Insurance Decisions in the Presence of Catastrophic Risks, IIASA Interim Report IR-97-068,(1997).

[7] T.Ermolieva, Yu.Ermoliev, V.Norkin, Spatial Stochastic Model for Optimization Capacity of Insurance Networks Under Dependent Catastrophic Risks: Numerical Experiments, IIASA Interim Report IR-97-028, (1997).

[8] Y.Ermoliev and V.Norkin, Stochastic generalized gradient method with application to insurance risk management, IIASA Interim Report IR-97-021, (1997), (Kibernetika i Sistemnyi Analiz. 2 (1998)).

[9] Y.Ermoliev and V.Norkin, Monte Carlo Optimization and Path Dependent Nonstationary Laws of Large Numbers, IIASA Interim Report IR-98-009, (1998).

[10] A.V.Fiacco and G.P.McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, J.Wiley & Sons, New York, 1968.

[11] V.V.Fyodorov, *Numerical Methods of Maximin*, Nauka, Moscow, 1979.

[12] Insurance Service Office, *The Impact of Catastrophes on Property Insurance*, New York, 1994.

[13] H.Konno, H.Yamazaki, Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market, Management Science 37 (1991), 519-531.

[14] G.MacDonald, *Persistance in Climate*, JSR-91-340, The MITRE Cooperation, McLean, Verginia, 22102-3481, 1992.

[15] H.M.Markowitz, *Mean Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Oxford, 1987.

[16] S.Messner, A.Golodnikov, A.Gritsevskii, A Stochastic Version of the Dynamic Linear Programming Model MESSAGE 3, Energy. vol.21, 9 (1996), 775-784.

[17] V.S.Mikchalevich , A.M.Gupal and V.I.Norkin, *Methods of Nonconvex Optimization*, Moscow, Nauka, 1987.

[18] V.Norkin, Generalized differentiable functions, *Kibernetika*. 1 (1980), 9-11 (In Russian, English translation in *Cybernetics*. vol. 16, 1).

[19] W.Ogryczaki, A.Ruszczynski, From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures, IIASA Interim Report 97-027, (1997).

[20] G.Pflug, Risk-Reshaping Contracts and Stochastic Optimization. IIASA, WP-96-142, (1996).

[21] E.Raik, Qualitative investigation of nonlinear stochastic programming problems, Izvestia Akademii Nauk Estonskoi SSR, Fizika i Matematika (Communications of the Estonian Academy of Sciences, Physics and Mathematics). vol. 21, 1 (1971), 8-14.

[22] R.J.-B.Wets, Challenges in stochastic programming, Math. Progr. vol.75 (1996), 115-135.